CERIF Tutorial

Jan Dvořák
May 29th, 2017
euroCRIS Spring Membership Meeting
Dublin, Ireland
Jan Dvořák
jan.dvorak@ff.cuni.cz

euroCRIS
• CERIF TG Leader since 2013
• CRIS 2012 (Prague, June 2012) Org. Committee Chair

Charles University in Prague, Faculty of Arts, Institute of Information Studies & Librarianship
• Researcher & Lecturer

Czech Technical University, Computing and Information Centre
• IS Analyst

InfoScience Praha
• Research, Development & Innovation Information System

This deck of slides is based on the CERIF Tutorial by Brigitte Jörg
CERIF TG Leader 2004-2012
What is Research Information?

The process of research
- Research projects
- Funding
- Research infrastructures

The research actors
- Researchers
- Institutions
- Funders
- Publishers
- Facility operators
- Associations

Research results
- Outputs (Publications, Research Datasets, Patents, …)
- Outcomes, Impacts
Who needs Research Information?

- Researchers: visibility, finding collaborations, competitors, CV generation
- Decision Makers: performance, strategic decisions, priorities, comparisons
- Project Managers: overview of ongoing activities
- Libraries: acquisition, dissemination
- Publishers: finding reviewers, editors
- Educators: integration of relevant findings into lectures and training
- Intermediaries / Brokers: finding research results of potential market or innovative value
- Enterprises: finding information for participation in projects, partnerships, usage of results
- Media: distribution and communication
- General Public: information and education, interest
- Research Organisations: integration and interoperability, strategic management
- Funding Organisations: distribution of programs, evaluation of results, finding reviewers

Research Information
Research Information Life-Cycle

- Monitoring
- Exchange
- Measurement

Store

Summarize

Disseminate
Common European Research Information Format

- CERIF is an EU Recommendation to Member States
- The European Commission (EC) has authorised euroCRIS to maintain and develop CERIF and its usage

http://cordis.europa.eu/cerif/
Model Levels

• **Conceptual Level** (Specification)
  Concepts relevant for the research domain and their relationships

• **Logical Level** (ER Model)
  Entities and their relationships

• **Semantic Layer** (Declared Semantics)
  A formalized controlled vocabulary describing a general contextual semantics of the research domain inline with the conceptual, logical and machine description
CERIF Base Entities

Person

Project

OrganisationUnit
CERIF Result Entities

ResultPublication

ResultPatent

ResultProduct
CERIF Result Entities

ResultPublication
- ID
- URI
- Title
- Subtitle
- Abstract
- Bibl. Note
- PublicationDate
- TotalPages
- StartPage
- EndPage
- Keywords

ResultPatent
- ID
- URI
- PatentNumber
- Title
- CountryCode
- RegistrationDate
- ApprovalDate
- Description
- Keywords

ResultProduct
- ID
- URI
CERIF Infrastructure Entities

- Facility
- Equipment
- Service
CERIF Infrastructure Entities

- **cfFacility**
  - cfName
  - cfDescription
  - cfKeywords
  - cfID
  - cfURI
  - cfAcronym

- **cfService**
  - cfName
  - cfDescription
  - cfKeywords
  - cfID
  - cfURI
  - cfAcronym

- **cfEquipment**
  - cfName
  - cfDescription
  - cfKeywords
  - cfID
  - cfURI
  - cfAcronym
Some CERIF Link Entities
Some CERIF Link Entities

Person_ResultPublication
  role=author

Person_OrganisationUnit
  role=research assistant

Person_Project
  role=principal investigator

Project_ResultPublication
  role=deliverable

OrganisationUnit_ResultPublication
  role=author's affiliation

Person

OrganisationUnit
  role=coordinator

Project

ResultPublication

Citation
CV
Prize
Qualification
ExpertiseAndSkills
Equipment
Facility
Funding
Service
ElectronicAddress
PostalAddress
Country
Currency
Language
Event
Metrics
Geographic
BoundingBox
A typical CERIF entity:

- Identifier
  - internal
- Attributes
  - the basic ones
- Multi-lingual attributes
- Classifications
  - Type
  - Status
  - Subject area
- Links
  - to other entities
  - recursive
Generic Linking Entity Structure

- Base object 1 (FK)
  - cfStartDate
  - cfEndDate
  - Time range of validity

- cfFraction (optional)
  - Fraction (optional)

- Base object 2 (FK)
  - role: cfClassification (FK)
Example: The Principal Investigator of project $P$ changes effective date $D$: $X$ is replaced by $Y$.

Before:

- Date range: $-\infty .. +\infty$
- Role: Principal Investigator, cfClassification
- $P \leftarrow X$

After:

- Date range: $-\infty .. D$
- Role: Principal Investigator, cfClassification
- $P \leftarrow X$
- Date range: $D .. +\infty$
- Role: Principal Investigator, cfClassification
- $Y$
Some CERIF Link Entities

**Unary classification:**
- Type
- Status
- Subject area

**Binary classifications:**
- Role

<table>
<thead>
<tr>
<th>Table 1: cfPers_Pers</th>
<th>Table 2: cfProj_OrgUnit</th>
</tr>
</thead>
<tbody>
<tr>
<td>cfPersId1</td>
<td>ID</td>
</tr>
<tr>
<td>cfPersId2</td>
<td>ID</td>
</tr>
<tr>
<td>cfClassId</td>
<td>ID</td>
</tr>
<tr>
<td>cfClassSchemId1</td>
<td>ID</td>
</tr>
<tr>
<td>cfFraction</td>
<td>Float</td>
</tr>
<tr>
<td>cfStartDate</td>
<td>Timestamp(6) NN (PK)</td>
</tr>
<tr>
<td>cfEndDate</td>
<td>Timestamp(6) NN (PK)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 3: cfOrgUnit_Equip</th>
</tr>
</thead>
<tbody>
<tr>
<td>cfOrgUnitId</td>
</tr>
<tr>
<td>cfEquipId</td>
</tr>
<tr>
<td>cfClassId</td>
</tr>
<tr>
<td>cfClassSchemId</td>
</tr>
<tr>
<td>cfFraction</td>
</tr>
<tr>
<td>cfStartDate</td>
</tr>
<tr>
<td>cfEndDate</td>
</tr>
<tr>
<td>cfCurrCode</td>
</tr>
<tr>
<td>cfAvailability</td>
</tr>
<tr>
<td>cfConditions</td>
</tr>
<tr>
<td>cfPrice</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 4: cfOrgUnit_Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>cfOrgUnitId</td>
</tr>
<tr>
<td>cfClassId</td>
</tr>
<tr>
<td>cfClassSchemId</td>
</tr>
<tr>
<td>cfFraction</td>
</tr>
<tr>
<td>cfStartDate</td>
</tr>
<tr>
<td>cfEndDate</td>
</tr>
<tr>
<td>cfCurrCode</td>
</tr>
<tr>
<td>cfAvailability</td>
</tr>
<tr>
<td>cfConditions</td>
</tr>
<tr>
<td>cfPrice</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 5: cfClass_Class</th>
</tr>
</thead>
<tbody>
<tr>
<td>cfClassId1</td>
</tr>
<tr>
<td>cfClassId2</td>
</tr>
<tr>
<td>cfClassSchemId1</td>
</tr>
<tr>
<td>cfClassSchemId2</td>
</tr>
<tr>
<td>cfFraction</td>
</tr>
<tr>
<td>cfStartDate</td>
</tr>
<tr>
<td>cfEndDate</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 6: cfProj_Pers</th>
</tr>
</thead>
<tbody>
<tr>
<td>cfProjId</td>
</tr>
<tr>
<td>cfPersId</td>
</tr>
<tr>
<td>cfClassId</td>
</tr>
<tr>
<td>cfClassSchemId</td>
</tr>
<tr>
<td>cfFraction</td>
</tr>
<tr>
<td>cfStartDate</td>
</tr>
<tr>
<td>cfEndDate</td>
</tr>
</tbody>
</table>
Measuring Impact in CERIF (MICE)

MICE, a JISC-funded Project coordinated by Richard Gartner, Kings College, London, UK
CERIF Measurement & Indicator

Is an Aggregation Entity

cfMeasureIdentifier
cfCountInteger
cfCountIntegerChange
cfCountFloatingPoint
cfCountFloatingPointChange
cfValueJudgementalNumeric
cfValueJudgementalNumericChange
cfValueJudgementalText
cfValueJudgementalTextChange
cfURI
Measurement & Indicator (some examples)

– economic and commercial
  • economic
    – impact on business
      » improving performance of existing businesses
      • increased turnover by 1.2M€ in 2012
      • time savings of 14.56%
      • reduced costs by 42%
    » new products/processes
      • creating numbers of new products/services
      • commercialising / other success measures

Extract from the MICE List of Indicators
CERIF 1.6
CERIF Semantic Layer

Allows to capture any Schema or Structure
- Flat Lists
- Thesauri
- Classification Systems (e.g. SKOS, ...)
- Taxonomies
- Ontologies

Open / Extensible in all directions
- New Schemas
- New Concepts / Terms
- New Relationships

Enables to manage
- Roles / Types Semantics
- Subject Headings
- Archiving (Time component)

Allows for Mappings between Schemes
CERIF Federated Identifiers

- **ResultPublication**
  - ISBN
  - ISSN
  - DOI
  - WoS Accession Number
  - Scopus EID
  - PubMed Central ID

- **Person**
  - Social Security Number
  - Staff Id in HR system
  - Author identifier
    - ORCID, IdRef, DAI, ResearcherID, ScopusID

- **Project/Grant**
  - Funder’s reference number
  - Organisation’s reference number

- **Organisation**
  - VAT Identification Number
  - FundRefID
  - GridID

- **Classification**
  - External Code
CERIF Federated Identifiers

- Records the “tag” by which an object is known elsewhere
- For any Base, Result, Infrastructure, or 2nd Level entity
- “Identifier Types” classification scheme
- (optionally) Connected to a Service representing the issuer of the identifier
  - Usually an information system
- For point-to-point interchange
- XML namespace
- XML Schema
  - Based on the ER model
<CERIF xmlns="urn:xmlns:org:eurocris:cerif-1.6-2">
  <cfProj>
    <cfProjId>internal-project-identifier</cfProjId>
    <cfAcro>ACRO</cfAcro>
    <cfURI>http://www.project-url.ac.uk/acro.html</cfURI>
    <cfTitle cfLangCode="en" cfTrans="o">The title of the project</cfTitle>
    <cfAbstr cfLangCode="en" cfTrans="o">The goals of the project</cfAbstr>
  </cfProj>
  <cfProj_Class>
    <cfClassId>infrastructure-project-uuid</cfClassId>
    <cfClassSchemeId>project-types-scheme-uuid</cfClassSchemeId>
  </cfProj_Class>
  <cfFedId>
    <cfFedId>PROJECT NUMBER</cfFedId>
    <cfClassId>project-number-uuid</cfClassId>
    <cfClassSchemeId>federated-identifier-type-uuid</cfClassSchemeId>
  </cfFedId>
  <cfProj_OrgUnit>
    <cfOrgUnitId>orgunit-1-identifier</cfOrgUnitId>
    <cfClassId>coordinator-uuid</cfClassId>
    <cfClassSchemeId>orgunit-project-roles-scheme-uuid</cfClassSchemeId>
    <cfStartDate>from-datetime</cfStartDate>
    <cfEndDate>to-datetime</cfEndDate>
  </cfProj_OrgUnit>
</CERIF>
CERIF 1.6 XML Interchange Format

XML Schema-based

Separate namespace

urn:xmlns:org:eurocris:cerif-1.6-2 for CERIF 1.6

Used in:

OpenAIRE Guidelines for CRIS managers 1.0

CERIF API specification (-> Arch TG)

euroCRIS CERIF CRIS Reference Implementation
Ongoing work: CERIF XML Update

• More readable XML
• Better connection with the research domain
• Reduce fragmentation

→ presentation
CERIF development

By the CERIF Task Group of euroCRIS

Adopting open-source software projects tools & best practise
CERIF highlights

- Right level of abstraction
- Normalized model
  - Record information only once
  - Reference rather than copy
- Versatile Semantic Layer
- Time-based relationships
- Clean design, regular structure

www.eurocris.org
Metadata Layers

- Discovery metadata
  - DC, VIVO, MODS, METS, eGMS, DCAT, ...

- Contextual metadata
  - CERIF

- Detailed metadata
  - Domain-specific standards

---

Generate

Reference
The CERIF Evolution

CERIF 2000 Model

- Data Model
- Multilinguality
- Controlled Vocabulary
- Roles / Types
- User-driven
- EC Recommendation to Member States

CERIF 2006 / 2008 Model

- Data Model
- Model Normalization
- Robust/Consistent Structure
- Extensible Structure
- Semantic Layer
- XML Exchange Specification
- Elaboration on Publication
- CERIF Core Semantics (2008 1.2)

Similar Ideas
UN/UNESCO
OECD
CODATA

EU Working Group on Research Databases Workshop

Acronym: ERGO
Participant: Keith Jeffery, Anne Asseron, many more
Organisations: Rutherford Appleton, University of Bergen, ...